MATE-T580 Practical Data Science using R
Assignment 4
Predicting Number of Received Applications by US Colleges

In this assignment, you will apply concepts related to the linear regression model on the college.csv dataset. The dataset is based on the US News publication that ranks US colleges. The specific task is to predict the number of applications received by a college based on the predictors in the dataset. You will approach the problem from three different angles:

- First, build a linear model based on all predictors in the dataset.
- Second, refine your model based on a combination of expert judgement/common sense and trial and error with respect to variable selection.
- Third, use Lasso regularization for feature selection.

In all three approaches, pay attention to how you validate the model and which performance metrics are used to assess the model performance.

Appendix

The college.csv file can be downloaded from the course github webpage:
https://github.com/maherharb/MATE-T580/tree/master/Datasets

Here's a quick guide to the data variables:

Private: A factor with levels No and Yes indicating private or public university

Apps: Number of applications received

Accept: Number of applications accepted

Enroll: Number of new students enrolled

Top10perc: Pct. new students from top 10% of H.S. class

Top25perc: Pct. new students from top 25% of H.S. class

F.Undergrad: Number of fulltime undergraduates

P.Undergrad: Number of parttime undergraduates

Outstate: Out-of-state tuition

Room.Board: Room and board costs

Books: Estimated book costs

Personal: Estimated personal spending

PhD: Pct. of faculty with Ph.D.'s

Terminal: Pct. of faculty with terminal degree

S.F.Ratio: Student/faculty ratio

perc.alumni: Pct. alumni who donate

Expend: Instructional expenditure per student

Grad.Rate: Graduation rate